

# **AMERICAN COMMUNITY SURVEY: HOUSING UNIT AND GROUP QUARTERS WEIGHTING**

Mark E Asiala  
Decennial Statistical Studies Division  
Bureau of the Census

For presentation at the October 18, 2007 Meeting  
Census Advisory Committee of Professional Associations

U S C E N S U S B U R E A U

## 11. WEIGHTING AND ESTIMATION

### A. OVERVIEW

Beginning in 2010, three sets of American Community Survey (ACS) estimates will be calculated annually for specified geographic areas, using data collected over three different time periods. In general, the ACS will produce and publish estimates for the same set of statistical, legal, and administrative entities as the previously published Census long form: the nation, states, American Indian and Alaska Native (AIAN) areas, counties (*municipios* in Puerto Rico), minor civil divisions (MCDs), incorporated places, and census tracts, among others (see Chapter XIII.B). Three sets of estimates will be published, with the total population of a geographic area determining which sets will be released.

- Multi-year estimates based on five calendar years of sample data will be published for all statistical, legal, and administrative entities, including census tracts, block groups, and small incorporated places, such as cities and towns. These estimates will be based on data collected during the 60 months of the five most recent collection years.
- For geographic entities with populations of at least 20,000, three-year estimates will be computed based on data collected during the 36 months of the three most recent collection years.
- For geographic entities with populations of at least 65,000, single-year estimates will be computed based on data collected during the 12 months of the most recent calendar year.

When subsequent three- and five-year period estimates are produced, data from the most recent year will replace data from the earliest year of the previous estimation period.

The basic estimation approach is a ratio estimation procedure<sup>32</sup> that results in the assignment of two sets of weights: a weight to each sample person record and a weight to each sample housing unit (HU) record. For any given tabulation area, a characteristic total is estimated by summing the weights assigned to the people, households, families, or HUs possessing the characteristic. Estimates of person characteristics are based on the person weight. Estimates of family, household, and HU characteristics are based on the HU weight. As with most household surveys, weights are used to bring the characteristics of the sample more into agreement with those of the full population by compensating for differences in sampling rates across areas, differences between the full sample and the interviewed sample, and differences between the sample and independent estimates of basic demographic characteristics (Alexander et al. 1997).

The next section describes the methodology for computing the 2005 ACS single-year estimates, which uses both HU and population data for people living in HUs. Later sections discuss the single-year weighting methodology for the total population—group quarters (GQs) and HU persons—which is planned for the 2006 ACS and beyond. The ratio estimation method also will be used to produce three- and five-year period estimates, and the general procedures for producing these are expected to be similar to the single-year period estimates.

Single-year weighting is implemented in three stages. In the first stage weights are computed to account for differential selection probabilities based on the sampling rates used to select the HU sample. In the second stage, weights of responding HUs are adjusted to account for nonresponding HUs. In the third stage, weights are controlled so that the weighted estimates of HUs and persons by age, sex, race, and Hispanic origin conform to estimates from the

---

<sup>32</sup> Ratio estimation is a method that takes advantage of auxiliary information (in this case, population estimates by sex, age, race, and Hispanic origin, and estimates of total housing units) to increase the precision of the estimates. This method is used to produce ACS estimates consistent with the population estimates from the Population Estimates Program (PEP) of the Census Bureau by these characteristics and the estimates of total housing units (HUs) for each county in the United States.

Population Estimates Program (PEP) of the Census Bureau at a specific point in time. The estimation methodology is implemented by “weighting area,” either a county or a group of less populous counties.

## **B. 2005 ACS WEIGHTING—PROBABILITY OF SELECTION**

The first stage of weighting involves two steps. In the first step, each HU is assigned a basic sampling weight, which is computed separately for the mail, computer-assisted telephone interviewing (CATI), and computer-assisted personal interviewing (CAPI) samples. In the second step, these sampling weights are adjusted to reduce variability in the monthly weighted totals.

### **a. Sampling Weight**

The first step is to compute the basic sampling weight for the HU based on the inverse of the probability of selection. This sampling weight is computed as a multiplication of the base weight ( $BW$ ) and a CAPI subsampling factor ( $SSF$ ). The base weight  $BW$  for an HU is calculated as the inverse of the final overall sampling rate as given in Chapter 4, Table 4-1. HUs sent to CAPI are eligible to be subsampled at one of the rates described in Table 4-2. Those selected for the CAPI subsample, and for which no late mail return is received in the CAPI month, are assigned a CAPI  $SSF$  equal to the inverse of their subsampling rate; those not selected for the CAPI subsample receive a factor of 0.0. HUs eligible for CAPI, and for which mail returns are received in the CAPI month, as well as HUs not eligible for CAPI (cases whose interviews were completed by mail or by CATI), receive a CAPI  $SSF$  of 1.0. A new weight—i.e., the weight after CAPI subsampling factor ( $WSSF$ ), equal to the base weight times the CAPI subsampling factor—is

calculated for every HU.<sup>33</sup> Table 11.1 summarizes computation of the *WSSF* by the weighting step and the type of sample.

TABLE 11.1  
COMPUTATION OF THE WEIGHT AFTER CAPI SUBSAMPLING FACTOR (*WSSF*)

Weighting step	Monthly sample				
	Mail respondent	CATI respondent	CAPI sampled units	CAPI non-sampled units	CAPI eligible, but then becomes mail respondent
(a) Base Weight ( <i>BW</i> )	1 / (overall sampling rate)	1 / (overall sampling rate)	1 / (overall sampling rate)	1 / (overall sampling rate)	1 / (overall sampling rate)
(b) CAPI Subsampling Factor ( <i>SSF</i> )	1	1	1 / (CAPI subsampling rate)	0	1
(c) Weight after Subsampling Factor ( <i>WSSF</i> )= (a) x (b)	1 / (overall sampling rate)	1 / (overall sampling rate)	1 / (overall sampling rate) x 1 / (CAPI subsampling rate)	0	1 / (overall sampling rate)

#### b. Variation in the Monthly Sample Factor

The goal of ACS estimation is to represent the characteristics in a geographic area across the specified time period. For single-year estimates, this period is 12 months, and for three- and five-year estimates, it is 36 and 60 months, respectively. The annual sample is allocated into monthly samples. The monthly sample becomes a basis for the operations of the ACS data collection, preparation, and processing, including weighting and estimation.

<sup>33</sup> After each of the subsequent weighting steps, with one exception that will be noted, a new weight is calculated as the product of the new factor and the weight following the previous step. For additional details about the weighting steps discussed in this and the following section, see Asiala (2004).

The data for HUs assigned to any sample month can be collected during a three-month time period. For example, the households in the January sample month can have their data collected in January, February, or March. Each HU in a sample belongs to a tabulation month (the month the case is completed). This is either the month given by a respondent on the first page of the completed mail questionnaire, or the month the interview is completed by CATI or CAPI.

Because of seasonal variations in response patterns, the number of HUs in tabulation months may vary, thereby over-representing some months and under-representing other months in the single- and multi-year estimates.<sup>34</sup> For this reason, an even distribution of HU weights by month is desirable. To smooth out the total weight for all sample months, a variation in monthly response factor (*VMS*) is calculated for each month<sup>35</sup> as:

*VMS* = (total sample base weights of all HUs in that sample month) / (total adjusted weight after CAPI subsampling factor of all HUs interviewed in that sample month)

$$= \frac{\sum_{j \in Month_i} B W_{ij}}{\sum_{j \in Month_i} W S S F_{ij}},$$

where

$BW_{ij}$  = base weight for sampled HU  $j$  within the  $i$ th month,

---

<sup>34</sup> ACS respondents are asked to provide answers corresponding to the month of data collection, not to the month of sample selection. Some survey items use the time of interview as the reference period for the questionnaire. Others use time references such as “last month,” “past 12 months,” “past 3 months,” “last week,” and “last 4 weeks.” Characteristics collected by ACS may be affected by seasonal variation. The use of the floating reference period requires the assumption that the samples from each month represent the same population and can introduce selection bias in each month’s sample when this is not so.

<sup>35</sup> This adjustment factor is computed within each ACS weighting area (either a county or a group of less populous counties). The index for weighting area is suppressed in this and all other formulas for weighting adjustment factors.

$WSSF_{ij}$  = adjusted HU weight after the CAPI subsampling factor for interviewed HU  $j$  within the  $i$ th month.

Table 11.2 illustrates the computation of the *VMS* adjustment factor within a particular county. In this example, if the total base weight (*BW*) within each month is given as 100 on line 1 of this table, and the total *WSSF* within each month is given on line 5, then the *VMS* is computed as the ratio of line 1 to line 5 for each month.

TABLE 11.2  
EXAMPLE OF COMPUTATION OF *VMS*

		Month				
		Mar	Apr	May	Jun	Jul
Line 1:	Total Base Weights ( <i>BW</i> ) Across Released Samples	100	100	100	100	100
	Total Weights After CAPI Subsampling ( <i>WSSF</i> ) by Mode:					
Line 2:	(a) Mail	55 (Mar sample)	45 (Apr sample)	40 (May sample)	45 (Jun sample)	50 (Jul sample)
Line 3:	(b) CATI	30 (Feb sample)	25 (Mar sample)	30 (Apr sample)	30 (May sample)	25 (Jun sample)
Line 4:	(c) CAPI	30 (Jan sample)	25 (Feb sample)	20 (Mar sample)	25 (Apr sample)	30 (May Sample)
Line 5:	Total Weights <i>WSSF</i> Across Modes (a+b+c)	115	95	90	100	105
Line 6:	<i>VMS</i> Adjustment Factor	$100 \div 115$	$100 \div 95$	$100 \div 90$	$100 \div 100$	$100 \div 105$

The adjusted weights after the variation of monthly response adjustment (*WVMS*) are a product of the weights after CAPI subsampling factor (*WSSF*) and the variation of monthly response factor (*VMS*). When the *VMS* factor is applied, the total *VMS* weights (*WVMS*) across

all HUs tabulated in a sample month will be equal to the total base weight of all HUs selected in that month's sample. The result is that each month contributes approximately 1/12 to the total single-year estimates. In other words, the single-year estimates of ACS characteristics are presented as a 12-month average that smoothes out any possible seasonal variation of monthly data. Analogously, each month contributes approximately 1/36 and 1/60 to the three- and five-year estimates, respectively.

Table 11.3 presents the five summary statistics of the *VMS* values for the year 2005 across all weighting areas within each month.

TABLE 11.3  
FIVE SUMMARY STATISTICS OF THE 2005 *VMS* FOR EACH MONTH

Statistic	Month											
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Minimum	0.521	0.644	0.670	0.686	0.630	0.669	0.682	0.640	0.644	0.624	0.645	0.635
Q1*	0.942	0.952	0.924	0.973	0.942	0.936	0.959	0.914	0.979	0.926	0.926	0.981
Median	1.010	1.019	0.991	1.041	1.009	0.999	1.039	0.977	1.060	0.996	0.993	1.060
Q3*	1.091	1.091	1.066	1.124	1.083	1.074	1.113	1.047	1.141	1.070	1.071	1.140
Maximum	5.097	3.203	2.027	2.339	2.543	2.726	2.810	1.650	3.223	1.750	2.087	2.316

\* Q1 and Q3 are the first and third quartiles, respectively.

## C. 2005 ACS WEIGHTING—NONINTERVIEW ADJUSTMENT

The noninterview adjustment uses three factors to account for sample HUs for which an interview is not completed. During data collection, nothing new is learned about the HU or person characteristics of noninterviewed HUs, so only characteristics known at the time of sampling can be used in adjusting for them. In other surveys and censuses, characteristics that have been shown to be related to HU response include census tract, building type (single- versus multi-unit structure), and month of data collection (Weidman et al. 1995). Within counties, if a



sufficient number of sample HUs are available to fill the cells of a three-way cross-classification table formed by these variables, we can simultaneously adjust for these three factors. Since tracts have fairly small sample sizes, however, and the sample is evenly spread across all 12 months of a year, usually this will not be the case. As a result, the noninterview adjustment is carried out in two steps—one based on building type and census tract, and one based on building type and tabulation month. Once these steps are completed and the factors are applied, the sum of the weights of the interviewed HUs will equal the sum of the VMS weights of the interviewed plus noninterviewed HUs.

Note that ineligible units such as vacant units and deletes<sup>36</sup> are excluded from the noninterview adjustment. The weight corresponding to these HUs remains unchanged during this stage of the weighting process since it is assumed that all vacant units or deletes are properly identified in the field and therefore are not eligible for the noninterview adjustment. The weighting adjustment is carried out only for the occupied, temporarily occupied, or noninterviewed HUs, so as to account for the eligible but noninterviewed HUs. After completion of the adjustment to the weights of the interviewed HUs, the noninterviewed HUs can be dropped from subsequent weighting steps; their assigned weights will be equal to 0.

The noninterview adjustment steps are applied to all HUs interviewed by any mode—mail, CATI, or CAPI respondents. However, nearly all noninterviewed HUs belong to the CAPI sample, so characteristics of CAPI nonrespondents may be closer to those of CAPI respondents than to mail and CATI respondents. To account for this possible mode-related noninterview bias,

---

<sup>36</sup> Deletes or out-of-scope addresses fall into three categories: (1) addresses of living quarters that have been demolished, condemned, or are uninhabitable because they are open to the elements; (2) addresses that do not exist; and (3) addresses that identify commercial establishments, units being used permanently for storage, or living arrangements known as group quarters.

a mode noninterview adjustment factor is computed after the two previously mentioned noninterview adjustment steps.

**a. Calculation of the First Noninterview Adjustment Factor**

In this step, all HUs are placed into adjustment cells based on the cross-classification of building type (single- versus multi-unit structures) and census tract. If a cell contains fewer than 10 interviewed HUs, it is collapsed with an adjoining<sup>37</sup> tract until the collapsed cell meets the minimum size of 10. Cells with no noninterviews are not collapsed, regardless of size, unless they are forced to collapse with a neighboring cell that fails the size criterion. The first noninterview adjustment factor (*NIFI*) for each eligible cell is:

*NIFI* = (total HU weight after variation in monthly response factor of interviewed occupied and temporarily occupied HUs and noninterviewed HUs) / (total HU weight after variation in monthly response factor of interviewed occupied and temporarily occupied HUs)

$$= \frac{\sum_{j \in \text{Resp}_i} WVMS_{ij} + \sum_{j \in \text{Nonresp}_i} WVMS_{ij}}{\sum_{j \in \text{Resp}_i} WVMS_{ij}},$$

where

$WVMS_{ij}$  = adjusted HU weight after the variation in monthly response adjustment for the  $j$ th HU within the  $i$ th adjustment cell

---

<sup>37</sup> Data are sorted by the weighting area, building type, and tract. Within a building type, a tract that has 10 or more responses is put in its own month. A tract that has no nonresponses and some responses (even though the total is fewer than 10) is put in its own month. A tract that has nonresponses and fewer than 10 responses is collapsed with the next tract. When determining the adjoining tract for collapsing, if the final tract needs to be collapsed, it will be collapsed with the previous, rather than the next tract.

All occupied and temporarily occupied interviewed HUs are adjusted by this first noninterview factor. Vacant and deleted HUs are assigned a factor of 1.0, and noninterviews are assigned a factor of 0.0. That is,

$$WNIF1_{ij} =$$

$$= WVMS_{ij} \times NIF1_i \quad \text{if } j\text{th unit within the } i\text{th adjustment cell is an interviewed} \\ \text{occupied or temporarily occupied HU,}$$

$$= WVMS_{ij} \quad \text{if the } j\text{th unit within the } i\text{th adjustment cell is a vacant or deleted} \\ \text{HU,}$$

$$= 0 \quad \text{if the } j\text{th unit within the } i\text{th adjustment cell is a noninterviewed} \\ \text{HU,}$$

where

$WNIF1_{ij}$  = adjusted HU weight after the first noninterview adjustment factor for the  $j$ th HU within the  $i$ th adjustment cell

#### **b. Calculation of the Second Noninterview Adjustment Factor**

The next step is the second noninterview adjustment. In this step, all HUs are placed into adjustment cells based on the cross-classification of building type and tabulation month. If a cell contains fewer than 10 interviewed HUs, it is collapsed with an adjoining<sup>38</sup> tabulation month

---

<sup>38</sup> Data are sorted by the weighting area, building type, and tabulation month. Within a building type, a tabulation month that has 10 or more responses is put in its own month. A tabulation month that has no nonresponses and some responses (even though the total is fewer than 10) is put in its own month. A tabulation month that has nonresponses and fewer than 10 responses is collapsed with the next tabulation month. When determining the adjoining month for collapsing, if the final tabulation month needs to be collapsed, it will be collapsed with the previous, rather than the next month. Check that footnote fits on same page. In printed copy, it was on two pages.

until the collapsed cell has at least 10 interviewed HUs. Cells with no noninterviews are not collapsed, regardless of size, unless they are forced to collapse with a neighboring cell that fails the size criterion. The second noninterview factor (*NIF2*) for each eligible cell is:

*NIF2* = (total HU weight after variation in monthly response factor of interviewed occupied and temporarily occupied HUs and noninterviewed HUs) / (total HU weight after first noninterview factor of interviewed occupied and temporarily occupied HUs)

$$= \frac{\sum_{j \in \text{Resp}_i} WVMS_{ij} + \sum_{j \in \text{Nonresp}_i} WVMS_{ij}}{\sum_{j \in \text{Resp}_i} WNIF1_{ij}} .$$

*NIF1* weights for all occupied and temporarily occupied interviewed HUs are adjusted by this second noninterview factor. Vacant and deleted HUs are given a factor of 1.0, and noninterviews are assigned a factor of 0.0. That is,

$$WNIF2_{ij} =$$

$$= WNIF1_{ij} \times NIF2_i \quad \text{if unit } j \text{ within adjustment cell } i \text{ is an interviewed occupied or temporarily occupied HU,}$$

$$= WNIF1_{ij} \quad \text{if the } j\text{th unit within the } i\text{th adjustment cell is a vacant or deleted HU,}$$

$$= 0 \quad \text{if the } j\text{th unit within the } i\text{th adjustment cell is a noninterviewed HU,}$$

where

*WNIF2<sub>ij</sub>* = adjusted HU weight after the second noninterview adjustment for the *j*th HU within the *i*th adjustment cell.

### c. Calculation of the Mode Bias Noninterview Factor

One element not accounted for by the two noninterview factors above is the systematic differences that exist between characteristics of households that return mail Census forms and those that do not (Weidman et al. 1995). The same element has been observed in the ACS across response modes. Virtually all noninterviews occur among the CAPI sample, and people in these HUs may have characteristics that are more similar to CAPI respondents than to mail and CATI respondents. Since the noninterview factors (*NIF1* and *NIF2*) are applied to all HUs interviewed by any mode, compensation may be needed for possible mode-related noninterview bias. The mode bias factor ensures that the total weights in the cells defined by a cross-classification of selected characteristics are the same as if the weight of noninterview HUs had been assigned only to CAPI HUs, but the factor distributes the weight across all respondents (within the cells) to reduce the effect on the variance of the resulting estimates.

The first step in the calculation of the mode bias noninterview factor (*MBF*) is to calculate an intermediate factor, referred to as the mode noninterview factor (*NIFM*). *NIFM* is not used directly to compute an adjusted weight; instead, it is used as a factor applied to the *WVMS* weight to allow the calculation of the *MBF*. The cross-classification cells are defined within weighting area (county or a group of less populous counties) by building type and tabulation month. Only HUs interviewed by CAPI and noninterviews are placed in the cells. If a cell contains fewer than 10 interviewed HUs, it is collapsed with an adjoining month. Cells with no noninterviews are never collapsed unless they are forced to collapse with a neighboring cell that fails the size criterion. The mode noninterview factor (*NIFM*) for a cell is:

$$NIFM = (\text{total HU weight after variation in monthly response factor of CAPI interviewed occupied and temporarily occupied HUs, and noninterviewed HUs}) / (\text{total HU weight after variation in monthly response factor of CAPI interviewed occupied and temporarily occupied HUs})$$

$$= \frac{\sum_{j \in \text{CAPI resp}_i} WVMS_{ij} + \sum_{j \in \text{Nonresp}_i} WVMS_{ij}}{\sum_{j \in \text{CAPI resp}_i} WVMS_{ij}}.$$

This mode noninterview factor is assigned to all CAPI interviewed occupied and temporarily occupied HUs. HUs for which interviews are completed by mail or CATI, vacant HUs, and deleted HUs are given a factor of 1.0. Noninterviews are given a factor of 0.0. The *NIFM* factor is used in the next step only. Note that the *NIFM* adjustment is applied to the *WVMS* weight rather than the HU weight after the first and second noninterview adjustments (*WNIF1* and *WNIF2*). That is,

$$WNIFM_{ij} =$$

$$= WVMS_{ij} \times NIFM_i \quad \text{if the } j\text{th unit within the } i\text{th adjustment cell is a CAPI interviewed occupied or temporarily occupied HU,}$$

$$= WVMS_{ij} \quad \text{if the } j\text{th unit within the } i\text{th adjustment cell is a mail or CATI completed interview, and vacant or deleted HU,}$$

$$= 0 \quad \text{if the } j\text{th unit within the } i\text{th adjustment cell is a noninterviewed HU,}$$

where

$WNIFM_{ij}$  = adjusted HU weight after the mode noninterview adjustment for the  $j$ th HU within the  $i$ th adjustment cell.

Next, a cross-classification table is defined within each weighting area by tenure (three categories: HU owned, rented, or temporarily occupied), tabulation month (twelve categories), and marital status of the householder (two categories: married/widowed, or single). All occupied

and temporarily occupied interviewed HUs are placed in their cells. If a cell has fewer than 10 interviewed HUs, the cells with the same tenure and month are collapsed across all marital statuses. If there are still fewer than 10 interviewed HUs, the cells with the same tenure are collapsed across all months. The mode bias factor (*MBF*) for each cell is then calculated as:

$MBF = (\text{total weight after mode noninterview factor of interviewed occupied and temporarily occupied HUs}) / (\text{total weight after second noninterview adjustment factor of interviewed occupied and temporarily occupied HU})$

$$= \frac{\sum_{j \in \text{Resp}_i} WNIFM_{ij}}{\sum_{j \in \text{Resp}_i} WNIF2_{ij}} .$$

All interviewed occupied and temporarily occupied HUs are adjusted by this mode bias factor, and the remaining HUs receive the factor 1.0. These adjustments are applied to the *WNIF2* weights. That is,

$$WMBF_{ij} =$$

$$= WNIF2_{ij} \times MBF_i \quad \text{if the } j\text{th unit within the } i\text{th adjustment cell is an interviewed occupied or temporarily occupied HU,}$$

$$= WNIF2_{ij} \quad \text{if the } j\text{th unit within the } i\text{th adjustment cell is a vacant or deleted HU, or noninterviewed HU,}$$

where

$WMBF_{ij}$  = adjusted HU weight after the mode bias factor adjustment for the  $j$ th HU within the  $i$ th adjustment cell.

#### **D. 2005 ACS WEIGHTING—HOUSING UNIT AND POPULATION CONTROLS**

This stage of weighting forces the ACS total HU and person weights to conform to estimates from the Census Bureau's Population Estimates Program (PEP). The PEP of the Census Bureau annually produces estimates of population by sex, age, race, and Hispanic origin; and total HUs for each county in the United States as of July 1. The ACS estimates are based on a probability sample, and will vary from their true population values due to sampling and nonsampling error (see Chapters 12 and 14). In addition, we can see from the formulas for the adjustment factors in the previous two sections that the ACS estimates also will vary based on the combination of interviewed and noninterviewed HUs in each tabulation month. As part of the process of calculating person weights for the ACS, estimates of totals by sex, age, race, and Hispanic origin are controlled to be equal to population estimates by weighting area. There are two reasons for this: (1) to reduce the variability of the ACS HU and person estimates, and (2) to reduce bias due to under-coverage of HUs and the people within them in household surveys. The bias that results from missing these HUs and people is partly corrected by using these controls (Alexander et al. 1997).

The assignment of final weights involves the calculation of three factors based on the use of HU and population controls. The first adjustment involves the use of independent HU estimates. A second and separate adjustment relies on the use of independent population estimates. The final adjustment is implemented to achieve consistency between the ACS HU estimates and the HU estimates from the PEP.

The estimates of total housing units for states and counties as of July 1 are produced by the U.S. Census Bureau on an annual basis. The estimates are computed based on a model:

$$\text{HU0X} = \text{HU00} + (\text{NC0X} + \text{NM0X}) - \text{HL0X}$$



where the suffix “X” indicates the year of the housing unit estimates, and

HU0X = Estimated 200X housing units

HU00 = Geographically updated Census 2000 housing units

NC0X = Estimated residential construction, April 1, 2000 to July 1, 200X

NM0X = Estimated new residential mobile home placements, April 1, 2000 to July 1, 200X

HL0X = Estimated residential housing loss, April 1, 2000 to July 1, 200X.

For more detailed background on the methodology used for the HU estimates, readers can visit

[http://www.census.gov/popest/topics/methodology/2004\\_hu\\_meth.html](http://www.census.gov/popest/topics/methodology/2004_hu_meth.html).

The population estimates, produced by the U.S. Census Bureau on an annual basis, are computed based on the following model:

$$P1 = P0 + B - D + NDM + NIM + NMM,$$

where

P1 = population at the end of the period

P0 = population at the beginning of the period

B = births during the period

D = deaths during the period

NDM = net domestic migration during the period

NIM = net international migration during the period

NMM = net military movement during the period.

For more detailed background on the methodology used for the population estimates, readers can visit [http://www.census.gov/popest/topics/methodology/2005\\_st\\_co\\_meth.html](http://www.census.gov/popest/topics/methodology/2005_st_co_meth.html).

Production of the population estimates for Puerto Rico is limited to population totals by *municipio*, and by sex-age distribution at the island level. For this reason, estimates of totals by

*municipio*, sex, and age for the PRCS are controlled so as to be equal to the population estimates. Currently, there are no HU controls available for Puerto Rico.

**a. Calculation of Housing Unit Post-Stratification Factor**

Note that both HU and population estimates used as controls have a reference date of July 1, which means that the 12-month average of ACS characteristics is controlled to the population with the reference date of July 1. If person weights are controlled to the population estimates as of that date, it is logical that HUs also are controlled to those estimates to achieve a consistent relationship between the two totals.

The housing unit post-stratification factor is employed to adjust the estimated number of ACS HUs by weighting area so as to agree with the PEP estimates. For the  $i$ th weighting area, this factor ( $HPF$ ) is:

$$HPF = (\text{PEP HU estimate}) / (\text{total adjusted HU weight after the mode bias factor of interviewed occupied, interviewed temporarily occupied, and vacant HUs})$$

$$= \frac{\widehat{HU}_i}{\sum_{j \in (\text{interviewed} + \text{vacant})} WMBF_{ij}}$$

where

$$\widehat{HU}_i = \text{PEP HU estimate for the } i\text{th weighting area.}$$

The denominator of the  $HPF$  formula aggregates the adjusted HU weight after the mode bias factor adjustment ( $WMBF$ ) across 12 months for the interviewed occupied, interviewed temporarily occupied, and vacant HUs. All HUs (including deleted HUs) then are adjusted by

this HU post-stratification factor, i.e.,  $WHPF = WMBF \times HPF$ , where  $WHPF$  is the adjusted HU weight after the HU post-stratification factor adjustment.

#### **b. Calculation of Person Weights**

The next step in the weighting process is to assign weights to persons via a three-dimensional raking-ratio estimation procedure. This is done so that (1) the combined estimates of spouses and unmarried partners conform to the combined estimate of married-couple and unmarried-partner households; (2) the estimate of householders conforms to the estimate of occupied housing units; and (3) the estimates for certain demographic groups are equal to their population estimates.

Each person in an interviewed occupied HU is assigned an initial person weight equal to the HU weight after the HU post-stratification factor is applied ( $WHPF$ ). Next, there are three steps of ratio adjustment. The first step uses three cells to classify persons by relationship to the householder (see Step I below). The second step uses two cells to classify persons by householder and non-householder. The third step uses up to 156 cells defined by race/Hispanic origin, sex, and age. The steps are defined as follows:

**Step I: Spouses and Unmarried Partners.** All persons are placed into one of three cells: (1) primary person in a two-partner relationship, (2) secondary person in a two-partner relationship, and (3) balance of population.

The first cell consists of all householders in a married-couple or unmarried-partner household. The second cell consists of all spouses or unmarried partners in those same households. All persons not fitting into cells (a) or (b) are put into the third cell. The marginals for the first two cells are both equal to the estimate of married-couple plus unmarried-partner households using the adjusted HU weight after the HU post-stratification factor is applied ( $WHPF$ ). The marginal for the third cell is equal to the PEP total population estimate minus the

sum of the marginals used for the other two cells. In this manner, the estimate of total population is controlled to the PEP total population estimate.

**Step II: Householders.** The second step assigns all persons to one of two cells: householders and non-householders. The marginal for householders is the estimate of occupied housing units using the *WHPF* weight. The marginal for non-householders is equal to the PEP total population estimate minus the marginal used for the first cell in order to control for total population.

**Step III: Race-Hispanic Origin/Sex/Age.** The third step assigns all persons to one of up to 156 cells: 6 classifications of race-Hispanic origin by sex by 13 age groups. The marginals for these rows at the weighting area level come from the PEP population estimates. Some weighting areas will not have sufficient sample to support all 156 cells, and in these cases some collapsing is necessary. This collapsing is done prior to the raking and remains fixed for all iterations of the raking.

Race and Hispanic origin are combined to define six unique race-ethnicity groups consistent with those used in weighting the Census 2000 long form. These groups are created by crossing “Non-Hispanic” with the five major single race groups, plus the group of all Hispanics regardless of race. The race-ethnicity groups are:

1. Non-Hispanic White
2. Non-Hispanic Black
3. Non-Hispanic American Indian and Alaskan Native
4. Non-Hispanic Asian
5. Non-Hispanic Native Hawaiian or Pacific Islander
6. Hispanic

The assignment of a single major race to a person can be complicated, because people can identify themselves as being of multiple races. People responding either with multiple races or “Other Race” are included in one of the six race-ethnicity groups for estimation purposes only. Subsequent ACS tabulations are based on the full set of responses to the race question.

Initial estimates of population totals are obtained from the ACS sample for each of the weighting race-ethnicity groups. These estimates are calculated based on the initial person weight of *WHPF*. Estimates from the Census Bureau’s PEP also are available for each weighting race-ethnicity group. These total population estimates are used to control ACS total population estimates to be equal to the PEP by weighting area.

The initial sample and population estimates are tested against a set of criteria which for each weighting race-ethnicity group require a minimum of 10 sample people, and that the ratio of the population control to the initial estimate is between (1/3.5) and 3.5.<sup>39</sup> If there are weighting race-ethnicity groups that do not satisfy these requirements, they are collapsed until all groups satisfy the collapsing criteria. Collapsing decisions are made following a specified order in the following way (see Asiala (2004) for further details.)

1. If the requirements are not met when all non-Hispanic race groups are combined then all weighting race-ethnicity groups are collapsed together and the collapsing is complete.
2. If the requirements are not met for Hispanics, the Hispanics are collapsed with the largest non-Hispanic non-White group.
3. If the requirements are not met for any non-Hispanic non-White group, it is collapsed with the largest (prior to collapsing) non-Hispanic non-White group.
4. If the largest collapsed non-Hispanic non-White group still does not meet the requirements, it is collapsed with the surviving non-Hispanic non-White groups in the following order until the requirements are met: Black, American Indian and Alaskan Native, Asian, and Native Hawaiian or Pacific Islander.

---

<sup>39</sup> This is done to reduce the effect of large weights on the variance of the estimates.

5. If the requirements are not met for the non-Hispanic White group, then it is collapsed with the largest non-Hispanic non-White group.

Within each collapsed weighting race-ethnicity group, the persons are placed in sex-age cells formed by crossing sex by 13 age categories: 0-4, 5-14, 15-17, 18-19, 20-24, 25-29, 30-34, 35-44, 45-49, 50-54, 55-64, 65-74, and 75+ years. If necessary, these cells also are collapsed to meet the requirements of the same sample size and a ratio between (1/3.5) and 3.5. The goals of the collapsing scheme are to keep children age 0-17 together whenever possible by first collapsing across sex within the first three age categories. In addition, the collapsing rules keep men age 18-54, women age 18-54, and seniors 55+ in separate groups by collapsing across age.

The initial sample cell estimates are then scaled and rescaled via iterative proportional fitting, or raking, so that the sum in each row or column consecutively agrees with the row or column household estimate (Steps I & II) or population estimate (Step III). This procedure is iterated a fixed number of times, and final person weights are assigned by applying an adjustment factor to the initial weights.

The scaling and rescaling between rows and columns is referred to as an iteration of raking. An iteration of raking consists of the following three steps. (The weighting matrix is included to facilitate the discussion below.) The three-step process has been split out into two tables, 11.4 and 11.5, for clarity.

TABLE 11.4

## STEPS I AND II OF THE WEIGHTING MATRIX

		Step II		Step I Control
		Householder	Non-Householder	
Step I	Householder in two-partner relationship			Survey estimate of married-couple and unmarried-partner households
	Spouse / unmarried partner in two-partner relationship			Same as above
	Balance of population			PEP total population est. minus the sum of the two controls above
	Step II Control	Survey estimate of occupied housing units	PEP total population est. minus the control for householders	

TABLE 11.5

## STEPS II AND III OF THE WEIGHTING MATRIX

			Step II		Step III Control
			Householder	Non-Householder	
Step III	Non-Hispanic White	0–4 Males			PEP population estimate for the collapsed cell by weighting area
		0–4 Females			
		...			
		75+ Females			
	Non-Hispanic AIAN	...			
	Non-Hispanic Asian	...			
	Non-Hispanic NHPI	...			
	Hispanic	...			
	Step II Control		Survey estimate of occupied housing units	PEP total population est. minus the control for householders	

**Step I.** At this step, the initial person weights are adjusted so the sum of the weights of householders in married-couple or unmarried-partner households and the sum of the weights of their spouses or unmarried partners are both equal to the survey estimate of married-couple and unmarried-partner households using the HU weight after the HU post-stratification factor

adjustment. The weights of all other persons are adjusted so that the sum of all weights is equal to the PEP total population estimate.

**Step II.** The Step I adjusted person weights are adjusted again so the sum of the weights of all householders is equal to the survey estimate of occupied housing units using the HU weight after the HU post-stratification factor adjustment. The Step I adjusted weights of all other persons are adjusted so that the sum of all weights is equal to the total population estimate.

**Step III.** The Step II adjusted person weights are adjusted a third time by the ratio of the population estimates of race-Hispanic origin/age/sex groups to the sum of the Step II weights for sample people in each of the demographic groups described previously.

The three steps of ratio adjustment are repeated in the order given above until the predefined stopping criterion is met. The stopping criterion is a function of the difference between Step II and Step III weights. The weights obtained from Step III of the final iteration are the final person weights.

A single factor, the person post-stratification factor (*PPSF*), is calculated at the person level, which captures the entire adjustment accomplished by the ratio-raking estimation. It is calculated as follows:

$$PPSF = (\text{final person weight}) / (\text{initial person weight})$$

The factor is calculated and applied to each person, so that their weights become the product of their initial weights and the factor.

Single-year estimates are produced for geographic areas with populations of at least 65,000, including incorporated places, for which population estimates also are published annually. Since population controls are applied at the weighting area level, occasionally the ACS estimate of total population for a large place within a weighting area may be far enough from its population



estimate to cause confusion among data users. To avoid these anomalies, methodologies are being investigated to control HU person weights to total population for places with populations of at least 65,000 within weighting areas.

### c. Calculation of Final Housing Unit Factors

Prior to the calculation of person weights, each HU has a single weight, and the sum over the year of these weights by weighting area is equal to the HU estimates. A new HU weight is computed by taking into account the characteristics of the householder in the HU. In each interviewed occupied HU, the householder defined as the reference person (one of the persons who rents or owns the HU) is identified. Adjustment of the HU weight to account for the householder characteristics is done by assigning a householder factor (*HHF*) for an HU equal to the person post-stratification factor (*PPSF*)<sup>40</sup> of the householder. Their *PPSF*s give an indication of under-coverage for households whose householders have the same demographic characteristics; applying the *HHF* adjusts for the resultant bias. Vacant HUs are given an *HHF* of 1.0 because they have no person weights.

The adjusted HU weight accounting for householder characteristics is computed as a multiplication of the adjusted HU weight after the HU post-stratification factor adjustment (*WHPF*) with the householder factor (*HHF*), i.e.,  $WHHF = WHPF \times HHF$ , where *WHHF* is the adjusted HU weight after the householder factor adjustment. The HU weight after the householder factor adjustment becomes the final HU weight.

This weighting procedure results in two separate sets of weights, one for HUs and one for persons residing within HUs. However, since the housing unit weight is equal to the person

---

<sup>40</sup> In the calculation of person weights, the *PPSF* is used to adjust person weight so that the ACS population estimates conform to IPE estimates by demographic characteristics.

weight of the householder, the survey will produce logically consistent estimates of occupied housing units, households, and householders. With this weighting procedure, the survey estimate of total housing units will differ slightly from the PEP total housing unit estimates. The difference nationally, however, is less than 0.1%.

The summary statistics of the final HU and person weights (respondents only) are given in Table 11.6.

TABLE 11.6  
FIVE SUMMARY STATISTICS OF THE FINAL HU AND PERSON WEIGHTS

Statistic	Housing Unit Weight	Person Weight
Minimum	3	2
Q1*	41	42
Median	53	54
Q3*	70	71
Maximum	781	865

\* Q1 and Q3 are the first and third quartiles, respectively.

## E. WEIGHTING FOR SINGLE-YEAR ESTIMATES OF TOTAL POPULATION

Estimates from the 2006 ACS will include data from both HUs and GQs, because people living in GQs are included in the sample for the 2006 data collection year. A project was undertaken to research an adequate method for single-year weighting including GQ residents while controlling household person estimates to PEP population estimates by demographic characteristics for weighting areas. The research compared four alternative options for controlling GQ persons, either separately or in combination with HU persons. The results showed that it is feasible to weight the GQ data at the state level by major GQ type group (see Table 11.7) and combine those results with the weighting of the household population by

weighting area to produce adequate estimates of the total population for all levels of aggregation. The choice of this methodology is further supported by the nature of the Population Estimate Program GQ population estimates which are updated and maintained by major GQ type.

The methodology for controlling HU and GQ person estimates will not affect the prior weighting steps. (These are described in the previous section for HUs.) Fewer steps are needed for GQs because all of their data are collected via personal visit, and the tabulation month is not used in any adjustments of GQ weights. The base weight is the reciprocal of the sampling rate and will be equal to 40 for all people sampled in GQs. In the small GQ strata, there are sometimes more than 15 people in a GQ. These situations occur when the GQ has more people than expected, which also applies to the large GQ stratum. In these cases, a subsample of the residents is taken so that only 10 are eligible for interview. The weight of these 10 people must be increased equally by a GQ subsampling factor to account for those not selected.

A noninterview factor is calculated to account for the GQ residents who do not complete an interview. This occurs in a single step where the noninterview cells are defined by combinations of GQ types, classified into major GQ type groups, by county. The major GQ type groups are shown in Table 11.7. If a cell contains fewer than 10 interviews and has any number of non-interviews, then cells are collapsed across counties but within state by major GQ type group. If the major GQ type group still fails, it is collapsed with other type groups in that state in the same institutional/non-institutional class as shown in the table.

TABLE 11.7  
MAJOR GQ TYPE GROUPS

Major GQ Type Group	Definition	Institutional/Non-Institutional
1	Correctional Institutions	Institutional
2	Juvenile Detention Facilities	Institutional
3	Nursing Homes	Institutional
4	Other Long-Term Care Facilities	Institutional
5	College Dormitories	Non-Institutional
6	Military Facilities	Non-Institutional
7	Other Non-Institutional Facilities	Non-Institutional

The GQ noninterview factor (*GQNF*) is similar to the noninterview factors applied to noninterviewed HUs, and weights the interviewed GQ persons in these adjustment cells up to the total weights of the interviewed and noninterviewed GQ persons by cell.

The third and last step in the GQ population weighting process is to apply the GQ person post-stratification factor (*GQPPSF*). This factor brings the ACS GQ person weights to conform to the PEP GQ population estimates for state by major GQ type group. The first step is to assign all GQ persons to one of seven major GQ types within the state. Like the household person weighting, there must be at least 10 GQ persons within the cell or it is collapsed with another major GQ type within the institutional/non-institutional classes. If further collapsing is necessary and all major GQ type groups within this dichotomy have been collapsed then all major GQ type groups are collapsed together into one cell for that state. In practice, the split between institutional and non-institutional is typically preserved as are many of the larger major GQ type groups.

## F. MULTI-YEAR ESTIMATES

Weighting for the combined HU and GQ multi-year estimates will be similar to that used to produce single-year estimates. The annual samples corresponding to the estimation period will

be combined, and all the weighting steps will be performed similarly to the single-year weighting. The main difference is that the ACS population and HU estimates are controlled to the means of the current PEP estimates for the multi-year estimation period. Note, however, that the resulting estimates may be different for the same weighting area under the one-, three-, and five-year estimation periods, especially for characteristics that change over time.

The weighting will be implemented based on the geographic definition of the weighting areas as of the last year of the estimation period; i.e., all period estimates will be produced for a consistent set of geographic areas. Additional details of multi-year estimation will be included in future versions of this document.